DOCUMENT RESUME

ED 402 327 TM 025 829

AUTHOR Bond, Linda A.
TITLE Norm-Referenced Testing and Criterion-Referenced Testing: The Differences in Purpose, Content, and Interpretation of Results.
INSTITUTION North Central Regional Educational Lab., Oak Brook, IL.
SPONS AGENCY Office of Educational Research and Improvement (ED), Washington, DC.
PUB DATE Aug 95
CONTRACT RP91002007
NOTE 9p.
AVAILABLE FROM North Central Regional Educational Laboratory, 1900 Spring Road, Suite 300, Oak Brook, IL 60521-1480.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Achievement Tests; *Criterion Referenced Tests; Educational Testing; *Norm Referenced Tests; Outcomes of Education; Scores; *Standardized Tests; State Programs; *Test Construction; Test Content; *Testing Programs; Test Use

ABSTRACT

Norm-referenced tests (NRT) help compare the performance of one student with the performances of a large group of students, while criterion-referenced tests (CRT) focus on "what test takers can do and what they know, not how they compare to others" (Anastasi, 1988). Both types of test can be standardized so that scores can be interpreted the same way for all students and schools. Test content for an NRT is selected according to how well it ranks students from high achievers to low, while the content of a CRT is selected by how well it matches the learning outcomes deemed most important, or on the basis of its importance in the curriculum. NRTs have come under attack recently because they tend to focus on low-level, basic skills. CRTs, on the other hand, give detailed information about how well a student has performed on each of the educational goals or outcomes included in the test. In 1994, 31 states administered NRTs and 33 administered CRTs, and 22 of these states administered both. Only two states rely on NRTs exclusively, and only one relies exclusively on a CRT. Most states also administer some other form of assessment. States will have to match their choice of assessment strategies to their intended purposes, the content they wish to assess, and the kinds of interpretation they want to make about student performance. (Contains six references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 402 327

Norm-Referenced Testing and Criterion-Referenced Testing: The Differences in Purpose, Content, and Interpretation of Results

NORTH CENTRAL REGIONAL EDUCATIONAL LABORATORY

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

By Linda A. Bond, Ph.D.
August, 1995

NCREL

BEST COPY AVAILABLE

TM025825

North Central Regional Educational Laboratory
1900 Spring Road, Suite 300
Oak Brook, IL 60521
(708)571-4700, Fax: (708) 571-4716

Jeri Nowakowski:	Executive Director
Deanna H. Durrett:	Director, RPIC
Lawrence B. Friedman:	Associate Director, RPIC
Linda Ann Bond:	Director of Assessment, RPIC
Lenaya Raack:	Editor

NCREL is one of ten federally supported educational laboratories in the country. It works with education professionals in a seven-state region to support restructuring to promote learning for all students—especially students most at risk of academic failure in rural and other schools.

The Regional Policy Information Center (RPIC) connects research and policy by providing federal, state, and local policymakers with research-based information on such topics as educational governance, teacher education, and student assessment policy.

© 1995 North Central Regional Educational Laboratory

This publication is based on work sponsored wholly or in part by the Office of Educational Research and Improvement (OERI), Department of Education, under Contract Number RP91002007. The content of this publication does not necessarily reflect the views of OERI, the Department of Education, or any other agency of the U.S. Government.

Norm-Referenced Testing and Criterion-Referenced Testing: The Differences in Purpose, Content, and Interpretation of Results

August 1995

by: Linda A. Bond, Ph.D.
Director of Assessment, Regional Policy Information Center
North Central Regional Educational Laboratory
1900 Spring Road, Suite 300
Oak Brook, Illinois 60521

Tests can be categorized into two major groups: norm-referenced tests and criterion-referenced tests. They differ in their intended purposes, the way content is selected, and the scoring process which defines how the test results must be interpreted. This brief paper will describe the differences between these two kinds of assessments and explain the most appropriate uses of each.

Intended Purposes

"Norm-referenced tests (NRTs) help compare one student's performance with the performances of a large group of students" (U.S. Congress, 1992, p. 168). A representative group of students, called the norm group, is given the test prior to its being sold to the public. Any student who then takes the test once it is published has his or her scores compared to those of the norm group and the student learns how he or she scored relative to the students who took the test when it was normed.

The norm group is usually a national sample of students. Tests such as the California Achievement Test (CTB/McGraw-Hill), the Iowa Test of Basic Skills (Riverside), and the Metropolitan Achievement Test (Psychological Corporation) are nationally normed in this way. Because norming a test is such an elaborate and expensive process, the norms are typically used by test publishers for seven years. All students who take the test during that seven-year period have their scores compared to the original norm group.

The major reason for using an NRT is to sort students. NRTs are designed to highlight achievement differences between and among students to produce a dependable rank order of students across a continuum of achievement from high achievers to low achievers (Stiggins, 1994). We might want to rank students in this way in order to place them in special remedial or gifted programs, or to select them for different ability level reading or mathematics instructional groups.

Criterion-referenced tests (CRTs), on the other hand, are focused on "what test takers can do and what they know, not how they compare to others" (Anastasi, 1988, p. 102). CRTs report how well students are doing relative to a predetermined performance level on a specified set of educational goals or outcomes included in the school, district, or state curriculum. Educators or policymakers may choose to use a CRT when they wish to see how well students have learned the knowledge and skills they are expected to learn. This information may be used as one piece of information to decide how well the student is learning the desired curriculum and how well the school is teaching that curriculum.

Both NRTs and CRTs can be standardized, meaning that we can compare the scores of one student or group of students against those of another. This means that we can assume that two students who receive the same score on the test have demonstrated the same level of performance. "A standardized test is one that uses uniform procedures for

administration and scoring in order to assure that the results from different people are comparable. Any kind of test—from multiple choice to essays to oral examinations—can be standardized if uniform scoring and administration are used" (U.S. Congress, 1992, p. 165). Most national, state, and district tests are standardized so that scores can be interpreted the same way for all students and schools.

Selection of Test Content

Another consideration for choosing an NRT or a CRT relates to the content of the test. Test content for an NRT is selected according to how well it ranks students from high achievers to low, while the content of a CRT is selected by how well it matches the learning outcomes deemed most important. While no test can measure everything of importance, the content selected for the CRT is selected on the basis of its importance in the curriculum while that of the NRT is selected by how well it discriminates among students.

NRTs have come under attack recently because they tend to focus on low-level, basic skills, which is in direct contrast to the emphasis on conceptual understanding and application of skills recommended by the latest research on teaching and learning. The National Council of Teachers of Mathematics has been particularly vocal about this concern. "A recent study of the six most commonly used commercial achievement tests found that at grade 8, on average, only 1 percent of the items were problem solving while 77 percent were computation or estimation" (Stenmark, 1991, p. 8). Since teachers tend to make sure they teach the content that is on the test, particularly if that test is used to judge how well they teach, they often emphasize low-level skills in the classroom (Corbett & Wilson, 1991). With both curriculum specialists and educational policymakers calling for more attention to higher level skills, these tests may be driving classroom practice in the opposite direction of reform.

Any national, state, or district test sends a message about what is important to learn and what level of performance is acceptable for students. Careful consideration of the content of the test that is selected or developed will therefore be an important decision.

Test Interpretation

As mentioned earlier, a student's performance on an NRT is interpreted in relation to the performance of a large group of similar students who took the test when it was first normed. For example, if a student receives a percentile rank score on the total test of 34, this means that he or she performed as well or better than 34 percent of the students in the norm group. This information is useful for deciding whether this student needs remedial assistance or is a candidate for a gifted program. However, it gives little information about what the student knows or can do, other than that he or she knows more of the test content than 34 percent of the students in the

norm group. Whether this is a good thing depends on whether the content of the NRT matches the knowledge and skills expected of those students. It is easier to ensure this match to expected skills with a CRT.

CRTs, on the other hand, give detailed information about how well a student has performed on each of the educational goals or outcomes included on that test. "For example, a CRT score might describe which arithmetic operations a student can perform or the level of reading difficulty he or she can comprehend" (U.S. Congress, 1992, p. 170). As long as the content of the test matches the content that is considered important to learn, the CRT gives the student, the teacher, and the parent more information about how much of the valued content has been learned than will an NRT.

Summary

Recognizing that public demands for accountability, and consequently for standardized tests, are not going to disappear, some states are designing tests that "reflect, insofar as possible, what we believe to be appropriate educational practice" (Stenmark, 1991, p. 9). In 1994, 31 states administered NRTs and 33 administered CRTs, and 22 of these states administered both. Only two states rely on NRTs exclusively and only one relies exclusively on a CRT. Most states also administer other forms of assessment, such as a writing sample, some form of open-ended performance assessment, or a portfolio (Council of Chief State School Officers, et al., 1994).

States will have to match their choice of assessment strategy(ies) to their intended purposes, the content they wish to assess, and the kinds of interpretations they wish to make about student performance. Once they have determined these three things, the choice becomes easier.

References

Anastasi, A. (1988). *Psychological testing*. New York: MacMillan Publishing Company.

Corbett, H. D., & Wilson, B. L. (1991). *Testing, reform and rebellion*. Norwood, NJ: Ablex Publishing Company.

Council of Chief State School Officers and North Central Regional Educational Laboratory. (1994). *State student assessment program database*. Oak Brook, IL: North Central Regional Educational Laboratory.

Stenmark, J. K. (Ed.) (1991). *Mathematics assessment: Myths, models, good questions, and practical suggestions*. Reston, VA: The National Council of Teachers of Mathematics (NCTM).

Stiggins, R. J. (1994). *Student-centered classroom assessment*. New York: Merrill, an imprint of Macmillan College Publishing Company.

U.S. Congress, Office of Technology Assessment. (1992). *Testing in America's schools: Asking the right questions* (OTA-SET-519). Washington, DC: U.S. Government Printing Office.

North Central Regional Educational Laboratory
1900 Spring Road, Suite 300
Oak Brook, IL 60521-1480
(708) 571-4700
Fax (708) 571-4716

BEST COPY AVAILABLE